Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting

Nguyen, Q., Rienties, B., Richardson, J.T.E. (2019). Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting. *Assessment and Evaluation in Higher Education*. DOI: 10.1080/02602938.2019.1679088.

Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting

Quan Nguyen[a]*, Bart Rienties[b], John T.E. Richardson[b]

[a]*School of Information, University of Michigan, United States*

[b]*Institute of Educational Technology, The Open University, Milton Keynes, UK*

*corresponding author. Email: quanngu@umich.edu

Notes on contributors:

Quan Nguyen is a Postdoctoral Research Fellow at the School of Information, University of Michigan. Quan was a PhD student at the Institute of Educational Technology at the Open University UK, funded by the Leverhulme Trust. He graduated as a Master (cum laude) of Economics and Information Management at Maastricht University. His research focuses on the use of learning analytics to inform and support learning design in online learning settings.

Bart Rienties is Professor of Learning Analytics at the Institute of Educational Technology at the Open University UK. He is programme director Learning Analytics within IET and head of Data Wranglers, whereby he leads of group of learning analytics academics who conduct evidence-based research and sense making of Big Data at the OU. His primary research interests are focussed on Learning Analytics, Computer-Supported Collaborative Learning, and the role of motivation in learning

John T. E. Richardson is emeritus professor in student learning at the UK Open University. He taught and researched in psychology at Brunel University from 1975 to 2001, when he moved to the Open University to take up a new chair in student learning and assessment in the Institute of Educational Technology. His main research interests were concerned with the relationship between students' perceptions of their courses of study in higher education and the approaches to studying that they adopt on those courses.

Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting

Abstract. Although the attainment gap between black and minority ethnic (BME) students and White students has persisted for decades, the potential causes of these disparities are highly debated. The emergence of learning analytics allows researchers to understand how students engage in learning activities based on their digital traces in a naturalistic setting. This study investigates the attainment gap by analysing the differences in behavioural engagement between different ethnic groups. Using multilevel models of academic performance, demographics, and online traces of 149,672 students enrolled in 401 modules in a distance learning setting, we confirmed the existing attainment gap. After controlling for other demographics, module characteristics and engagement, BME students were between 19% and 79% less likely to complete, pass, or achieve an excellent grade compared to White students. Given the same academic performance, BME students spent 4-12% more time on studying than White students. While the attainment gap remained persistent after controlling for academic engagement, our study further highlighted the inequality of attainment between BME and White students.

Keywords: attainment gap, learning analytics, distance learning, BME students, engagement.

Introduction

Decades of research into academic attainment in the UK have established a wide disparity in academic performance between White and Black and ethnic minority (BME) students (Connor, Tyers, Modood, & Hillage, 2004; Broecke & Nicholls, 2007; Richardson, 2008, 2015, 2018). The latest review by Richardson (2018) which synthesised data from multiple sources over the last 20 years showed that the odds of obtaining a good degree (i.e. first-class or upper second-class) in BME students are about half those in White students. This under-attainment effect was stronger for Black students compared to Asian students. The attainment gap in ethnicity persisted even after controlling for other factors such as age, gender, prior qualifications, quality of feedback, and self-reported engagement (Richardson, 2010, 2011; Richardson, Alden Rivers, & Whitelock, 2015). Nonetheless, there is much unknown about the causes of the attainment gap and what can be done to address this issue.

In the last eight years, the emergence of the interdisciplinary field of learning analytics has demonstrated its potential to identify students who may need additional support from an early stage and provide real-time interventions (Ferguson, 2012). By capturing and analysing fine-grained digital traces of online learning activities, researchers can gain an in-depth understanding of what, when, and how students engage with their study (Tempelaar, Rienties, & Giesbers, 2015; Rienties & Toetenel, 2016; Nguyen, Rienties, Toetenel, Ferguson, & Whitelock, 2017). Although there are numerous studies in learning analytics focusing on retention issues, attention has thus far not been paid to the use of learning analytics to address the attainment gap between White and BME students. This study investigates equity of attainment on a large scale in a distance learning setting by examining differences in the online behavioural engagement of 149,672 students enrolled in 401 online modules in the academic year 2017–2018.

The under-attainment of ethnic minority students

Black and ethnic minority is a broad term to describe a range of minorities living in the UK (Connor et al., 2004). BME students in our context primarily refer to students with Asian, African, or Caribbean origins who are considered as home students. At the time of university registration, students are asked to self-identify themselves from a list of different ethnic groups which is similar to that used in the UK national census. The most common index of attainment in previous research is based on the classification of first degrees (e.g. Bachelor). In the UK, a first-class or upper second-class honour is considered a ‘good’ degree.

Whether or not academic success can be measured and is relevant, can of course be debated (Sanders & Rose-Adams, 2014). Furthermore, as highlighted by Singh (2011) the simple BME-white categorisation should be critically debated, as the term ethnicity is socially constructed, and people may have multiple identities. For this paper, given the largely empirical focus using learning analytics we assume that academic success and ethnicity can objectively be measured, but elsewhere we have indicated the complex, fluid, and non-linear complexities of labelling and academic success (Rienties, Johan, & Jindal-Snape, 2015; Rogaten & Rienties, 2018).

The attainment gap between BME and White students has been well-documented in the literature. Connor (1996) surveyed 136 students graduated in 1993 and reported that BME graduates were less likely to obtain a first or upper second-class honours degree than Whites. The under-attainment issue was stronger in Black students than Indian or Chinese students. This pattern was confirmed again in subsequent studies (Connor et al., 2004; Broecke & Nicholls, 2007; Richardson, 2008, 2011, 2012, 2015; Richardson et al., 2015; Richardson, 2018).

The search for potential causes of the under-attainment issue in BME students has also been going on for decades. Firstly, entry qualifications as proxies of academic ability were attributed to the disparities in academic attainment. Richardson (2008) reported an increase in odds ratio in Black students from 0.33 to 0.60, and in Asian students from 0.50 to 0.71 after controlling for entry qualifications. That means the attainment gap can be explained by differences in student prior qualifications, with more BME students had a lower prior qualification than White students when they entered universities. However, the attainment gap persisted after controlling for other demographic and institutional variables (Broecke & Nicholls, 2007). Therefore, entry qualifications only explained half of the attainment gap.

Secondly, differences in students' experience were suspected to widen the attainment gap. Osler (1999) argued that BME students might encounter discriminatory or unconscious bias in teaching and assessment practices, which resulted in lower grades. However, there are mixed evidences to support this claim. Connor et al. (2004) survey on 1,300 undergraduate students followed up by 30 interviews in 29 different higher education institutions found that academic progress of BME students was influenced by many personal factors such as social isolation, finance, and time management. Nonetheless, these factors apply to all ethnic groups. Experience of direct discrimination did not seem to be directly associated with academic practices, but rather due to the lack of diversity of the student population, staff population, and local population. However, the effect of (unconscious) discrimination can make BME students feel isolated in group work, or emotionally stressful while integrating into the local environment. This might have an indirect effect on their academic performance. However, little attention has been paid to how these difficulties had hindered BME students' progress.

Furthermore, Richardson et al. (2015) examined the possible unconscious bias in the nature of feedback that 470 BME students and 470 matched White students received on their assignment from their tutors. However, given the same grade, there was no significant differences between the kinds of feedback received by BME and White students. In addition, the attainment gap remains consistent in both face to face and distance learning, which suggested that the nature of interactions with tutors was not attributable to the under-performance of BME students (Richardson, 2010).

Thirdly, differences in academic engagement between BME and White students have also been studied. Surveys on academic engagement conducted in the US, Netherlands, and the UK indicated that there was little difference between the two groups with respect to their engagement, and the attainment gap persisted after controlling for engagement variables (Severiens, ten Dam, & Blom, 2006; Severiens & Wolff, 2008; Richardson, 2011). An exception was the study in the US by Johnson, Crosnoe, and Elder (2001) which analysed self-report instruments on academic attachment and engagement of 8,104 middle school students and 2,482 high school students. The authors found that African American students were more engaged than were White and Hispanic American students. However, the use of self-report instruments to represent engagement has many limitations including response bias, sample bias, and failure to account for the inter-temporal aspects of engagement (i.e. a process happening over time than a fixed trait) (Richardson, 2004).

As discussed below, the introduction of learning analytics has opened up a new venue to investigate academic engagement in a naturalistic setting and take into account the dynamics of what, when, and how students engage in learning activities.

Learning analytics and academic engagement

In the last eight years, learning analytics has attracted a lot of attention from practitioners, managers, and researchers in education by shedding light on a large amount of (potentially) valuable data in education. Furthermore, learning analytics may provide the means to empirically test the validity and reliability of existing psychometric instruments and pedagogical theories in large-scale, naturalistic environments (Ferguson, 2012; Sclater, Peasgood, & Mullan, 2016). In early September 2018, the Joint Information Systems Committee (JISC) launched the world's first national learning analytics service for UK's higher education institions, with more than 30 universities signed up. Learning analytics has the potential to support students' academic success by offering in-time and personalised feedback. There are many examples of institutional learning analytics initiatives to support student retention across the globe such as OU Analyse at the Open University UK (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015), Course Signal at Purdue University (Arnold & Pistilli, 2012), or ECoach at the University of Michigan (Lonn, McKay, & Teasley, 2017). As an interdisciplinary field, learning analytics explores various sources of data such as institutional data (i.e. demographics, performance, course design), behavioural data (i.e. log-files, videos, discussion forums), psychological data (i.e. surveys, interviews, think aloud) and physiological data (i.e. eye-tracking, functional magnetic resonance imaging, heart rates, electroencephalogram). Learning analytics researchers have employed a wide range of analytical techniques such as social network analysis, inferential statistics, data mining, and machine learning.

Numerous studies have established the strong correlations between how students engage in online learning activities and their academic performance. For example, Tempelaar et al. (2015) study on 922 undergraduate students indicated that up 39%-51% of the variance in academic performance can be explained by behavioural traces. Nguyen, Huptych, and Rienties (2018) showed that high-performing students not only studied harder (i.e. spent more time on task) but also smarter (i.e. spent more time studying in advance) than low-performing students. Therefore, the disparities in academic attainment could be potentially linked to the level of engagement of students between ethnic groups.

However, the conceptualisation and measurement of engagement are rather complex (Sinatra, Heddy, & Lombardi, 2015; D'Mello, Dieterle, & Duckworth, 2017). Engagement can be viewed as a multi-dimensional construct ranging from behavioural engagement, emotional engagement, cognitive engagement, and agentic engagement (Sinatra et al., 2015). Sinatra et al. (2015) recommended that engagement should be considered on a continuum from person-centred to context-centred orientation. Towards the person-oriented direction, measurements of engagement consist of trace data, or physiological indicators such as eye-tracking, and heart rates. The focus of our study is behavioural engagement, as proxied by trace data in a virtual learning environment (VLE) in a distance learning setting.

The level of behavioural engagement is also mediated by instructional context. For example, two large-scale studies at the Open University UK on 111,256 students and 72,377 students respectively showed that a large proportion of the variance of students' behaviours is influenced by course characteristics such as learning design (Rienties & Toetenel, 2016; Nguyen et al., 2017). Similar studies in other contexts also reached the same conclusion (Gašević, Dawson, Rogers, & Gasevic, 2016). Therefore, this study aims to explain the attainment gap between BME and White students by controlling for behavioural engagement of students and heterogeneity across different instructional conditions.

Research questions

Firstly, we aim to confirm the established attainment gap while controlling for other demographics (i.e., age, gender, ethnicity, prior educational level, occupational status), module characteristics (i.e., module level of study, number of credits), and behavioural engagement (i.e., as measured by the time spent on the VLE).

1. What are the differences in academic attainment between BME and White students in a distance learning setting after controlling for other demographic factors, module characteristics, and behavioural engagement on the VLE?

Secondly, we explore the differences in behavioural engagement on the VLE across different ethnic groups, while controlling for other demographic factors and module characteristics.

2. What are the differences in behavioural engagement between BME and White students in a distance learning setting after controlling for other demographic factors, module characteristics, and academic outcome?

Methods

Setting and participants

The Open University is a distance-learning institution established in 1969 and the largest university in the UK. The OU's mission is to make education accessible to everyone regardless of their background. For most undergraduate modules, no formal qualifications are required. All modules are delivered in an online or in a blended format. The Open University has a distinctive population of students. Most of them are from the UK, with White ethnicity (87.45%), 26-45 years old, working full-time or part-time and have various prior educational levels (Rienties & Toetenel, 2016; Nguyen et al., 2017). This study included 149,672 students and their daily trace data in the VLE in 401 modules in the academic year 2017–2018.

Measurements

Demographics: Age, gender, ethnicity, prior educational level, occupational status.

The categorisation of ethnicity at the OU is similar to most other higher education institutions in the UK which included the following categories:

1. White; White - British; White - Irish; White - Scottish; Irish Traveller; Other White background.
2. Black or Black British - Caribbean; Black or Black British - African; Other Black background.
3. Asian or Asian British - Indian; Asian or Asian British - Pakistani; Asian or Asian British - Bangladeshi; Chinese; Other Asian background.
4. Mixed-White and Black Caribbean; Mixed-White and Black African; Mixed-White and Asian; Other mixed background.
5. Other ethnic background.

Since the OU has an open entry policy, there are a diverse population of students with different prior qualifications which were categorised into: No Formal Qualifications, Less than A Levels, A levels or equivalent, Higher education Qualifications, Post-graduate Qualifications, and Not known.

Academic attainment: While previous studies (Connor et al., 2004; Broecke & Nicholls, 2007; Richardson, 2008, 2011, 2012, 2015; Richardson et al., 2015; Richardson, 2018) have focused on the achievement of a 'first degree' as an index for academic attainment, this study focused on attainment at a module level. This allows us to examine engagement at a fine-grained level using trace data of students in their respective modules, and account for the differences of teaching and assessment practice across different modules. To achieve a comprehensive understanding of academic attainment, three measurements were used. Firstly, students were differentiated between those who completed the module and those who did not complete the relevant module. Not-completed students were defined as those who formally de-registered from the module or did not achieve any grades. Secondly, students who completed the module were differentiated by their average score of both continuous assessment scores and the final assessment/exam score. A student can either pass (average score ≥ 40) or fail (average score < 40). Thirdly, students who passed the module were differentiated between excellent (average score ≥ 75) and pass (40 ≤ average score < 75).

Module characteristics: Module level and number of credits reflect the level of difficulty (e.g. level 1=introductory, level 2=intermediate, level 3=advanced, and a postgraduate level). The number of credits represent the amount of expected total workload. Each credit equates one hour of studying. Most modules have 30 or 60 credits where full-time study equates to 120 credits per year. Modules with 0, 10, and 15 credits were excluded because they were short training modules. There are some access modules which introduce students to distance learning and university-level study, which were also excluded from the analysis.

Behavioural engagement: As a distance learning institution, the OU UK organises most of its learning activities online via a VLE. Therefore, we used the duration (in hours) on the VLE per day, aggregated per student per course as proxies of academic engagement in a distance learning context. Duration is defined as the time difference between two clicks. As pointed out by previous research (Kovanovic, Gašević, Dawson, Joksimovic, & Baker, 2016), this metric could be problematic due to outliers problems: 1) the inability to differentiate between active time and non-active time (students leave the respective web page open and go for a coffee), and (2) the last click of the day is followed by a click next day), which makes the duration excessively long. We processed outliers using an inter-quartile range of the cohort in each module. Outliers were not removed but cut off at the upper inter-quartile range of the respective module. The rationale for this method is that outliers can be attributed to both individual and instructional factors. For example, an excessively long duration could be due to a) students left their laptop opened, and b) certain modules require a higher workload. Therefore, using a combined inter-quartile range of individuals' time spent on the VLE in the same cohort, we can establish what is 'normal' for both the individual student and the module that they enrolled. We aggregated daily time spent on the VLE per student per module as follows:

- For completed students: From the course start date until the final assessment deadline.

- For non-completed students: From the course start date until the last date of registration.

Data analysis

To account for the hierarchical nature of our dataset (e.g. students are nested within modules), we will use a multi-level modelling to allow for random variance between modules (Goldstein, 2011). Firstly, three multilevel logistic regression models were fitted with academic attainment as binary dependent variables, predicting the likelihood of completing a module, passing a module, and achieving excellent grades respectively. Secondly, three multilevel linear regression models were fitted to predict the duration spent on the VLE. Independent variables were consequently added to each model, starting with ethnicity, followed by other demographics and module characteristics, and academic outcome. Since the duration on the VLE was positively skewed, a log-transformation was performed to satisfy the assumptions of normality in multilevel modelling. Diagnostic plots of models' residuals were examined which ensured that there were no severe violations of parametric assumptions (e.g. homogeneity of variances, multicollinearity, normal distribution). All the tests were carried out in R studio statistical software (v1.1.423) (R Core Team, 2016). The mixed effect logistic model was carried out using the glmer function in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Given our large sample size in both RQ1 and RQ2, we chose a more conservative cut-off significant value of 0.01 instead of 0.05 to mitigate the errors rate of detecting significant effect due to random chance in a large dataset (Lin, Lucas, & Shmueli, 2013). We also reported odds ratio with a 95% confidence interval to support readers' interpretation.

Results

Differences in academic attainment between BME and White students in a distance learning setting

Table 1 presents the descriptive statistics of average grades and duration spent on the VLE disaggregated by ethnicity and performance. The attainment gap was clear. Students with Asian (M=60.85, SD= 18.96) and Black (M=53.93, SD=18.59) ethnicity had lower average scores than White students (M=65.65, SD=19.09).

Table 1: Descriptive statistics of duration on the VLE and average grades

	N	Mean	SD	Min	Max
VLE duration per module (hours)					
By ethnicity					
Asian	5,997	45.03	56.90	0.00	785.86
Black	5,544	46.62	56.74	0.00	631.42
Mixed	3,863	43.70	54.39	0.00	809.15
Other	1,521	45.03	55.10	0.00	580.72
Refused	3,067	51.05	69.82	0.00	1,425.33
White	129,680	47.35	58.71	0.00	2,898.41
By performance					
Excellent	39,218	73.06	74.95	0.00	2,898.41
Pass	66,387	50.50	53.16	0.00	1,273.81
Fail	11,136	20.52	28.29	0.00	556.431
Not completed	32,931	18.73	33.51	0.00	1,425.33
Total	149,672	47.19	58.68	0.00	2,898.41
Average grades					
By ethnicity					
Asian	4,650	60.85	18.96	1	100
Black	4,100	53.93	18.59	1	100
Mixed	2,846	61.32	19.86	1	100
Other	1,174	60.38	20.33	1	100
Refused	2,405	65.25	19.74	1	100
White	101,566	65.65	19.09	1	100
By performance					
Excellent	39,218	83.18	6.06	75	100
Pass	66,387	61.32	9.18	40	74.5
Fail	11,136	21.69	11.77	1	39.5
Total	116,741	64.89	19.26	1	100

To control for other demographic factors and the heterogeneity between different modules, three multilevel logistic regressions were fitted (Table 2). The first model predicted the odds of completing a module. Students from a Black and Mixed ethnicity background had 19% and 22% lower chance of completing a module compared to White students respectively ($p<0.001$). However, there was no significant difference in the odds of completing a module between Asian and White students. The attainment gap got wider in model 2 and model 3. Compared to White students, students with Asian, Black, Mixed, and other ethnicity backgrounds were 25%, 60%, 31%, 44% less likely to pass module respectively (p<0.001). In model 3, students from with Asian, Black, Mixed, and other ethnicity backgrounds were 48%, 79%, 31%, 46% less likely to achieve an excellent grade respectively compared to White students (p<0.001).

Other controlled variables also influenced academic attainment. There were significant differences in academic attainment between occupational status. Students who were unable to work due to sickness or disability, or those who were unemployed, had a lower chance of completing, passing, and achieving excellent grades. Students who work a part-time job were 9% less likely to pass a module compared to students with a full-time job. However, there was no difference in the odds of completing or achieving excellent grades between students with part-time jobs and full-time jobs. Female students were more likely to complete (OR = 1.37, p<0.01) and achieve excellent grade (OR=1.05, p<0.01) than male students, although there was no difference in the likelihood of passing a module.

There was a negative association between age and academic attainment. Younger students were more likely to complete a course. Students with higher prior educational qualifications had a higher chance of completing, passing, and achieving excellent grades. There was no difference in the likelihood of completing a module between different levels of study. However, students who enrolled in level 2 modules and postgraduate level modules were 2.73 and 2.20 times more likely to pass than level 1 modules. However, students in postgraduate modules were 77% less likely to achieve an excellent grade compared to level 1 modules. Students who enrolled in 30-credit modules had a higher chance of completing and achieving an excellent grade compared to 60-credit modules. Finally, given the same group of students, 1% increase in the time spent on the VLE was associated with 2.15% – 3.10% increase in academic attainment. The Variance Partition Coefficient (VPC) indicated that 24.2% - 48.4% of the variance in individual academic attainment could be explained by module characteristics, which reinforced our decision of using a multilevel modelling approach. To sum up, our results confirmed the existing attainment gap which indicated that after controlling for other demographic factors, module characteristics and time spent on the VLE, BME students were 19-79% less likely to either complete, pass, or achieve an excellent grade compared to White students (Figure 1).

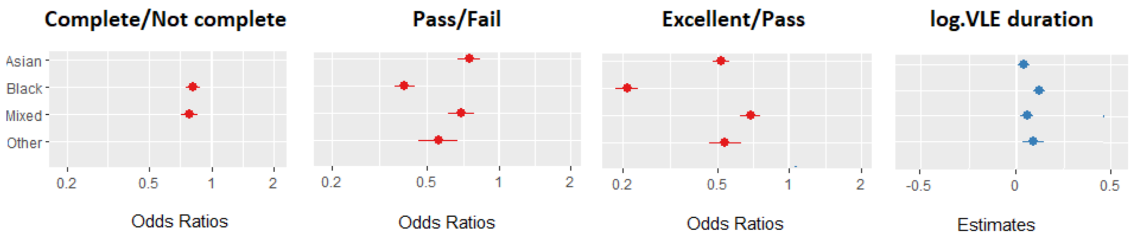

Figure 1: A summary of the differences in academic attainment and behavioural engagement

between different ethnic groups (White = reference category) of 149,672 students in 401 distance learning modules. All models have been controlled for age, gender, occupational status, prior qualifications, course level, number of credits, VLE duration (for modelling attainment), and academic performance (for modelling VLE duration). All $p<0.01$

Table 2: Multilevel logistic regression of academic attainment with random intercepts

	Completed/Not Completed		Pass/Fail		Excellent/Pass	
	Odds Ratio	CI	Odds Ratio	CI	Odds Ratio	CI
Fixed Parts						
(Intercept)	0.09 **	0.06 – 0.14	0.18 **	0.13 – 0.25	0.03 **	0.02 – 0.03
Ethnicity (ref=White)						
Asian	0.99	0.92 – 1.07	0.75 **	0.67 – 0.83	0.52 **	0.48 – 0.56
Black	0.81 **	0.75 – 0.88	0.40 **	0.36 – 0.44	0.21 **	0.19 – 0.23
Mixed	0.78 **	0.71 – 0.86	0.69 **	0.61 – 0.79	0.69 **	0.63 – 0.76
Other	0.92	0.79 – 1.07	0.56 **	0.46 – 0.67	0.54 **	0.46 – 0.63
Refused	0.96	0.86 – 1.07	0.84	0.72 – 0.99	0.81 **	0.74 – 0.90
Occupation status (ref = In full-time work)						
Doing unpaid voluntary work	0.87	0.75 – 1.02	0.63 **	0.51 – 0.77	1.06	0.92 – 1.23
In part-time work/self-employed	0.95	0.91 – 0.99	0.91 *	0.86 – 0.97	1.05	1.01 – 1.09
Information Refused	1.06	0.96 – 1.18	1.01	0.87 – 1.17	1.18 **	1.08 – 1.30
Looking after the home/family	0.76 **	0.72 – 0.81	0.73 **	0.67 – 0.80	1.09 *	1.03 – 1.15
Not in paid work for some other reason	1.06	0.97 – 1.15	1.34 **	1.17 – 1.53	1.41 **	1.30 – 1.53
Not Known	1.39 **	1.24 – 1.56	0.84	0.73 – 0.98	0.97	0.87 – 1.07
Retired from paid work	0.87	0.76 – 0.98	1.07	0.84 – 1.36	1.15	1.03 – 1.27
Unable to work: long-term sickness/disability	0.56 **	0.52 – 0.61	0.62 **	0.56 – 0.70	0.81 **	0.74 – 0.88

Unemployed and looking for a job	0.90 *	0.84 – 0.97	0.60 **	0.55 – 0.66	0.76 **	0.70 – 0.81
Gender (ref=Male)						
Female	1.37 **	1.32 – 1.42	1.03	0.98 – 1.09	1.05 *	1.02 – 1.09
Age (ref=35-46)						
Under 25	1.52 **	1.45 – 1.60	1.02	0.95 – 1.09	0.85 **	0.81 – 0.88
26-35	1.19 **	1.14 – 1.24	1.03	0.96 – 1.10	0.98	0.95 – 1.02
46-55	0.77 **	0.73 – 0.82	0.95	0.87 – 1.04	1.00	0.95 – 1.05
56 and over	0.63 **	0.58 – 0.69	0.74 **	0.64 – 0.85	0.80 **	0.74 – 0.86
Prior qualification (ref=Less than A levels)						
No Formal Qualifications	0.91	0.82 – 1.01	0.81 *	0.69 – 0.94	0.92	0.82 – 1.04
A Levels or equivalent	1.29 **	1.24 – 1.34	1.46 **	1.37 – 1.54	1.40 **	1.35 – 1.46
HE Qualification	1.22 **	1.17 – 1.28	1.44 **	1.35 – 1.53	1.74 **	1.67 – 1.82
PG Qualification	1.57 **	1.44 – 1.70	2.42 **	2.10 – 2.79	3.61 **	3.36 – 3.88
Not known	1.44 **	1.36 – 1.53	1.79 **	1.65 – 1.94	1.94 **	1.84 – 2.04
Course level (ref = OU Level 1)						
OU Level 2	0.75	0.45 – 1.26	1.43	0.96 – 2.13	0.69	0.51 – 0.95
OU Level 3	1.24	0.74 – 2.08	2.73 **	1.84 – 4.05	0.73	0.54 – 0.99
Postgraduate	0.62	0.38 – 1.02	2.20 **	1.47 – 3.30	0.23 **	0.17 – 0.31
Credits (ref = 60 credits)						
30 credits	2.00 **	1.39 – 2.87	1.05	0.79 – 1.40	1.74 **	1.40 – 2.18
log.VLE duration	3.10 **	3.05 – 3.14	3.39 **	3.31 – 3.47	2.15 **	2.11 – 2.19
Random Parts						

N groups	401	383	383
VPC groups	0.484	0.336	0.242
Observations	149672	116741	105605
Notes	** p<0.01 ** p<0.001, two-tailed*		

Differences in behavioural engagement between BME and White students in a distance learning setting

The average time spent on the VLE (hours) per module by Asian students (M=45.03, SD=56.90) and Black students (M=46.62, SD=56.74) was less than White students (M=47.35, SD=58.71) (Table 1). These differences are relatively small given the large standard deviations. In line with previous work (Nguyen et al., 2018), students with excellent performance (M=73.06, SD=74.95) had the highest time spent on the VLE, followed by passed students (M=50.50, SD=53.16), failed students (M=20.52, 28.29), and non-completed students (M=18,73, SD=33.51).

Table 3 reports three multilevel linear regression models. Model 1 predicted the duration spent on the VLE based on ethnicity. Model 2 controlled for other demographic and module characteristics. Model 3 accounted for academic performance. In model 1 and model 2, BME students spent 6-13% less time on the VLE than White students. However, when we controlled for academic performance, BME students spent 6-12% more time on the VLE than White students, to achieve the same performance. Compared to students with a full-time job, students with a part-time job spent 2% more time on the VLE, students who were retired spent 22% more time on the VLE, while students who were unable to work or unemployed spent 4-17% less time on the VLE. Female students spent 15% less time on the VLE than male students. Older students spent more time on the VLE than younger students. Compared to students who passed, excellent students spent 44% more time on the VLE whereas students who failed and did not complete the module spent 117% to 174% less time on the VLE. Module heterogeneity explained from 24.9% to 39.3% of the variance in time spent on the VLE between individual students. Model comparison using AIC showed that model 3 showed the best fit.

Table 3: Multilevel linear regression of time spent on Virtual Learning Environment with random intercepts

	log.VLE duration		log.VLE duration		log.VLE duration	
	B	CI	B	CI	B	CI
Fixed Parts						
(Intercept)	3.11 **	3.03 – 3.19	3.35 **	3.16 – 3.54	3.77 **	3.57 – 3.97
Ethnicity (ref=White)						
Asian	-0.08 **	-0.12 – -0.05	-0.03	-0.07 – 0.00	0.04 *	0.01 – 0.07
Black	-0.07 **	-0.10 – -0.03	-0.09 **	-0.13 – -0.06	0.12 **	0.09 – 0.15
Mixed	-0.13 **	-0.17 – -0.08	-0.06 *	-0.10 – -0.01	0.06 *	0.02 – 0.09
Other	-0.05	-0.12 – 0.02	0.00	-0.07 – 0.07	0.09 *	0.03 – 0.14
Refused	0.05	-0.00 – 0.10	0.03	-0.02 – 0.08	0.05	0.01 – 0.09
Occupation status (ref = In full-time work)						
Doing unpaid voluntary work			0.01	-0.06 – 0.08	0.07	0.02 – 0.13
In part-time work/self-employed			0.01	-0.00 – 0.03	0.02 *	0.01 – 0.04
Information Refused			0.02	-0.02 – 0.07	0.00	-0.03 – 0.04
Looking after the home/family			-0.08 **	-0.11 – -0.06	0.02	-0.00 – 0.04
Not in paid work for some other reason			-0.07 *	-0.10 – -0.03	-0.07 **	-0.10 – -0.04
Not Known			-0.21 **	-0.26 – -0.16	-0.17 **	-0.20 – -0.13
Retired from paid work			0.31 **	0.25 – 0.37	0.22 **	0.18 – 0.27
Unable to work: long-term sickness/disability			-0.25 **	-0.29 – -0.21	0.02	-0.01 – 0.05

Unemployed and looking for a job	-0.19 **	-0.23 – -0.16	-0.04 *	-0.07 – -0.01
Gender (ref=Male)				
Female	-0.14 **	-0.16 – -0.13	-0.15 **	-0.16 – -0.14
Age (ref=35-46)				
Under 25	-0.35 **	-0.37 – -0.33	-0.30 **	-0.32 – -0.29
26-35	-0.22 **	-0.24 – -0.20	-0.18 **	-0.19 – -0.16
46-55	0.20 **	0.18 – 0.23	0.19 **	0.17 – 0.21
56 and over	0.32 **	0.29 – 0.36	0.33 **	0.30 – 0.36
Prior qualification (ref=Less than A levels)				
No Formal Qualifications	-0.05	-0.11 – -0.00	0.02	-0.02 – 0.06
A Levels or equivalent	0.05 **	0.03 – 0.06	-0.06 **	-0.08 – -0.05
HE Qualification	0.13 **	0.11 – 0.15	-0.02	-0.03 – -0.00
PG Qualification	0.14 **	0.11 – 0.18	-0.13 **	-0.16 – -0.10
Not known	0.21 **	0.19 – 0.24	-0.02	-0.04 – 0.00
Course level (ref = OU Level 1)				
OU Level 2	-0.00	-0.24 – 0.23	0.06	-0.20 – 0.32
OU Level 3	-0.11	-0.34 – 0.13	-0.14	-0.39 – 0.12
Postgraduate	-0.06	-0.28 – 0.17	0.13	-0.12 – 0.37
Credits (ref = 60 credits)				
30 credits	-0.23 *	-0.39 – -0.07	-0.35 **	-0.52 – -0.17
Outcome (ref=Pass)				
Excellent			0.44 **	0.43 – 0.45

Fail			-1.17 **	-1.19 – -1.15
Withdrawn			-1.74 **	-1.75 – -1.72
Random Parts				
N groups	401	401		401
VPC groups	0.249	0.258		0.393
Observations	149672	149672		149672
R^2 / Ω_0^2	.163 / .163	.190 / .190		.471 / .471
AIC	520927.757	516109.514		452482.39
Notes	* *p<0.01* ** *p<0.001 , two-tailed*			

Discussion

In terms of RQ1, in line with previous research (Richardson, 2008, 2018), our findings confirmed the attainment gap between BME and White students in a distance learning setting after controlling for other demographic factors, module characteristics, and behavioural engagement on the VLE. Black students had an odds ratio from 0.21 to 0.81 to either complete, pass, or achieve an excellent grade compared to their White peers. Asian students had an odds ratio from 0.52 to 0.75 compared to White students. Mixed ethnicity groups had an odds ratio from 0.69 to 0.78. Other ethnicities had an odds ratio from 0.54 to 0.56. In other words, holding everything else constant (i.e. same prior qualifications, the same level of academic engagement), BME students still had significantly lower chance to complete, pass, or achieve an excellent grade compared to White students. The cause of under-attainment of BME students is indicated to lies elsewhere.

In terms of RQ2, our second finding further emphasised this inequality whereby on average, BME students spent more time on the VLE than White students to achieve the same level of performance. Black students spent 12% more time on the VLE than White students, when statistically controlling for variations in outcome variables. Asian, Mixed, and Other ethnicities spent 4%, 6%, 9% more time on the VLE than White students respectively. Despite being the most disadvantaged, Black students had the highest level of engagement than all other ethnic groups.

In contrast, when controlling for academic outcome, White students spent the least time on the VLE compared to other ethnicities. Our findings contradict previous studies which found no significance in self-reported engagement level between BME and White students (Severiens et al., 2006; Severiens & Wolff, 2008; Richardson, 2011) except the study of Johnson et al. (2001). A possible explanation lies in the differences in how engagement was measured. In our study, we employed a learning analytics approach which removed the biases of self-report and may provide a more accurate representation of engagement in a distance learning setting. At the same time, as mentioned before, the measures of behavioural engagement cannot capture other dimensions of engagement such as cognitive, emotional and agentic engagement. On the one hand, a higher level of behavioural engagement could reflect a higher level of effort. On the other hand, a longer time spent on the VLE could be a sign of struggling or confusion during the learning process. This could be due to a lack of familiarity with the curricum, lack of self-regulation skills, a difference in access to and usage of technology and IT, or a hidden systematic bias in teaching and assessment practices that may require more effort from BME students. Regardless of the underlying reasons, we have empirically confirmed that the attainment gap cannot be attributable to the differences in the level of engagement between BME and White students.

This study has important implications for institutions in tackling the under-attainment issue. Future interventions should focus on developing study skills and early-detecting struggling signals of BME and other students. For example, future studies could focus on the timing of engagement (i.e. studying in advance or catching up) of BME students compared to their White peers (Nguyen et al., 2018). An increasing number of hours spent on catching on previous materials could be a struggling signal for some groups of students, while for others this might indicate strong engagement. Moreover, learning analytics study at a micro level could be able to identify which study materials are causing delays in the learning progress of BME students.

Although this study involved a large number of students, readers should be aware of its caveats and generalisability. Firstly, trace data as proxies of behavioural engagement can only capture online activities. These proxies are more representative of the total engagement in a distance online learning than face-to-face learning. However, the time students spent on studying off-line was not taken into account in this study. Secondly, while we can capture what, when, and how much students engaged in learning activities, we do not know why or the underlying learning strategies of students. Last but not least, this study was conducted at a distance learning institution with an open entry policy focussing on attainment within a module, with a diverse population of students from a broad range of age, prior qualifications, and occupations. Whether the attainment gap had a long-term impact on academic success in follow-up modules and the qualification as a whole will need to be explored in future research.

Conclusion

This study investigated the attainment gap between BME students and White students by examining the differences in behavioural engagement in a distance learning setting. Using multilevel models on daily trace data of 149,672 students who enrolled in 401 modules in the academic year 2017–2018, we found that after controlling for other demographics, module characteristics and engagement, BME students were 19-79% less likely to complete, pass, or achieve an excellent grade compared to White students. Subsequent analyses indicated that BME students would have to put more effort than White students to achieve the same level of outcome. BME students spent 4-12% more time on studying than White students. Our study further highlighted the inequality of attainment in ethnic minority groups.

Acknowledgements

To be added

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

To be added

References

Arnold, K. E., & Pistilli, M. D. (2012). *Course signals at Purdue: using learning analytics to increase student success.* In proceedings of the Proceedings of the 2nd international conference on learning analytics and knowledge. 267-270

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Broecke, S., & Nicholls, T. (2007). *Ethnicity and degree attainment*.Vol., No. Department for Education and Skills. http://https//www.education.gov.uk/publications/eOrderingDownload/RW92.pdf

Connor, H. (1996). *Ethnic Minority Graduates: Differences by Degrees* (1851842357).Vol. Report 309, No. University of Sussex, Institute for Employment Studies. . Brighton. Retrieved from: https://www.employment-studies.co.uk/system/files/resources/files/309.pdf

Connor, H., Tyers, C., Modood, T., & Hillage, J. (2004). *Why the difference? A closer look at higher education minority ethnic students and graduates*.Vol. Research report 552, No. Institute for Employment Studies London. Retrieved from: http://www.bristol.ac.uk/media-library/sites/ethnicity/migrated/documents/educationreport.pdf

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning. *Educational psychologist*, *52*(2), 104-123.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5-6), 304-317.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28,* 68-84.

Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922): John Wiley & Sons.

Johnson, M. K., Crosnoe, R., & Elder, G. H. (2001). Students' Attachment and Academic Engagement: The Role of Race and Ethnicity. *Sociology of Education, 74*(4), 318-340.

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2016). Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. *2016, 2*(3), 81-110.

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., & Wolff, A. (2015). OU Analyse: analysing at-risk students at The Open University. *Learning Analytics Review*, 1-16.

Lin, M., Lucas, H. C., Jr, & Shmueli, G. (2013). Research commentary—too big to fail: large samples and the p-value problem. *Information Systems Research, 24*(4), 906-917.

Lonn, S., McKay, T. A., & Teasley, S. D. (2017). Cultivating Institutional Capacities for Learning Analytics. *New Directions for Higher Education, 2017*(179), 53-63.

Nguyen, Q., Huptych, M., & Rienties, B. (2018). *Linking students' timing of engagement to learning design and academic performance.* In proceedings of the Eighth International Learning Analytics & Knowledge Conference, Sydney, Australia. 141-150

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior, 76*, 703-714.

Osler, A. (1999). The Educational Experiences and Career Aspirations of Black and Ethnic Minority Undergraduates. *Race Ethnicity and Education, 2*(1), 39-58.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Richardson, J. T. E. (2004). Methodological Issues in Questionnaire-Based Research on Student Learning in Higher Education. *Educational Psychology Review, 16*(4), 347-358.

Richardson, J. T. E. (2008). The attainment of ethnic minority students in UK higher education. *Studies in Higher Education, 33*(1), 33-48.

Richardson, J. T. E. (2010). Face-to-face versus online tuition: Preference, performance and pass rates in white and ethnic minority students. *British Journal of Educational Technology, 43*(1), 17-27.

Richardson, J. T. E. (2011). The academic engagement of White and ethnic minority students in distance education. *Educational Psychology, 31*(2), 123-139.

Richardson, J. T. E. (2012). The attainment of White and ethnic minority students in distance education. *Assessment & Evaluation in Higher Education, 37*(4), 393-408.

Richardson, J. T. E. (2015). The under-attainment of ethnic minority students in UK higher education: what we know and what we don't know. *Journal of Further and Higher Education, 39*(2), 278-291.

Richardson, J. T. E. (2018). Understanding the Under-Attainment of Ethnic Minority Students in UK Higher Education: The Known Knowns and the Known Unknowns. In J. Arday & H. S. Mirza (Eds.), *Dismantling Race in Higher Education: Racism, Whiteness and Decolonising the Academy* (pp. 87-102). London: Palgrave Macmillan.

Richardson, J. T. E., Alden Rivers, B., & Whitelock, D. (2015). The role of feedback in the under-attainment of ethnic minority students: evidence from distance education. *Assessment & Evaluation in Higher Education*, *40*(4), 557-573.

Rienties, B., Johan, N., & Jindal-Snape, D. (2015). Bridge building potential in cross-cultural learning: a mixed method study. *Asia Pacific Education Review, 16*(1), 37-48.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333-341.

Rogaten, J., & Rienties, B. (2018). Which first-year students are making most learning gains in STEM subjects? *Higher Education Pedagogies, 3*(1), 161-172.

Sanders, J., & Rose-Adams, J. (2014). Black and minority ethnic student attainment: A survey of research and exploration of the importance of teacher and student expectations. *Widening Participation and Lifelong Learning, 16*(2), 5-27.

Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education*.Vol. 8, No.: https://www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v3.pdf

Severiens, S., ten Dam, G., & Blom, S. (2006). Comparison of Dutch ethnic minority and majority engineering students: social and academic integration. *International Journal of Inclusive Education*, *10*(1), 75-89.

Severiens, S., & Wolff, R. (2008). A comparison of ethnic minority and majority students: social and academic integration, and quality of learning. *Studies in Higher Education, 33*(3), 253-266.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational psychologist, 50*(1), 1-13.

Singh, G. (2011). *Black and minority ethnic (BME) students' participation in higher education: improving retention and success. A synthesis of research evidence*.Vol., No. Higher Education Academy. https://www.heacademy.ac.uk/system/files/bme_summit_final_report.pdf

Tempelaar, D., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167.